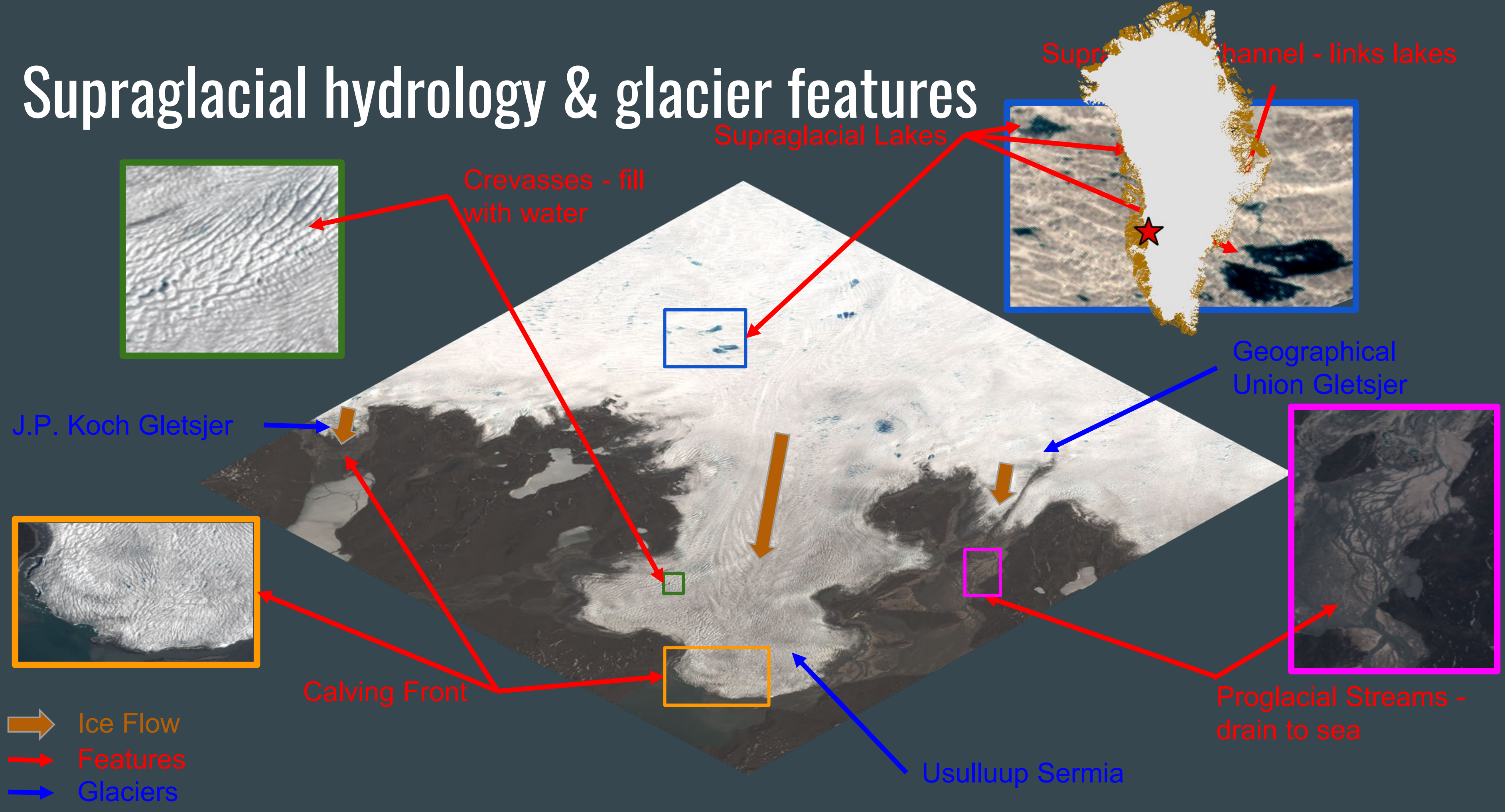
Towards Automated Mapping of Supraglacial Hydrology Dynamics Within an Ice Sheet Digital Twin.

Lessons learned from the 4D Antarctica and 4D Greenland Studies.

Diarmuid Corr, Mal McMillan, Ce Zhang, Amber Leeson & Emily Glen

Supraglacial hydrology & glacier features



Importance supraglacial hydrology

Increased runoff.	• Higher global temperature increases meltwater production. • More meltwater - more runoff - more sea level rise.
Injection of meltwater to the bed (in Greenland*).	• Lake drainage, by hydrofracture, introduces water to bed. • Reduces basal friction which may increase ice flow velocity. • Surface water stored in lakes, modulates flow of water to the bed.
Ice-shelf fracture (in Antarctica*).	• Rapid drainage suggested as mechanism for break-up of ice shelves - e.g. Larsen B. • Increases ice discharge from upstream glaciers.
Increasing albedo.	• Darker coloured water decreases ice surface albedo - increases absorption of incoming solar energy. • Potential positive feedback - enhances local melting.
Cryo-hydrologic warming.	• Heat transferred through the passage of meltwater to the bed. • Affects englacial and subglacial thermal conditions.

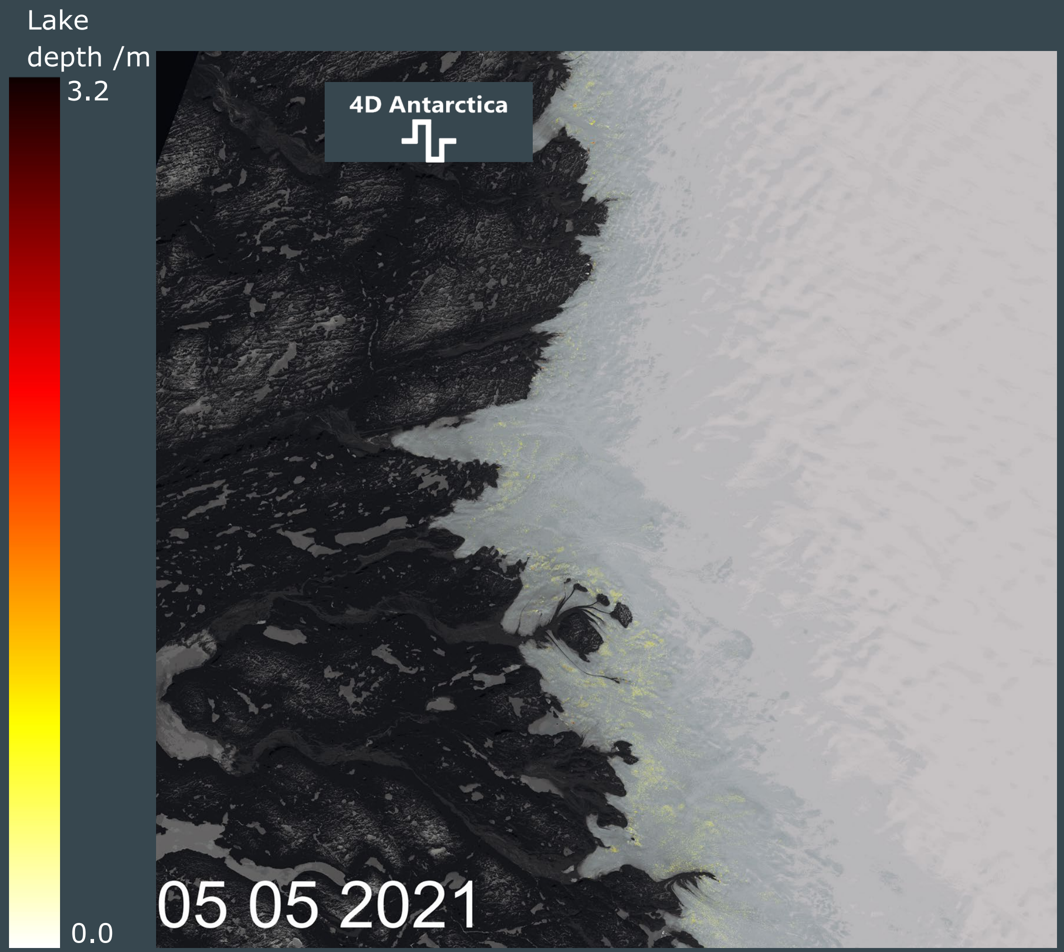
Typical Melt-Season

- 10,000s features in each melt-season.

- Feature distribution and characteristics vary throughout melt-season.

- Surface melt extent peaks in:
  - July/August for Greenland.
  - January for Antarctica.

- Antarctic melt-season typically shorter than Greenlandic.



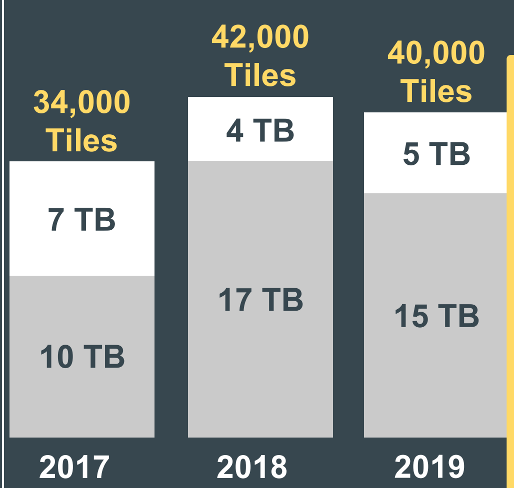
Large data quantities

Antarctica

Greenland

Mean: 1 Tile ≃ 0.5 GigaBytes

Sentinel-2: Revisit times of 5-10 days.



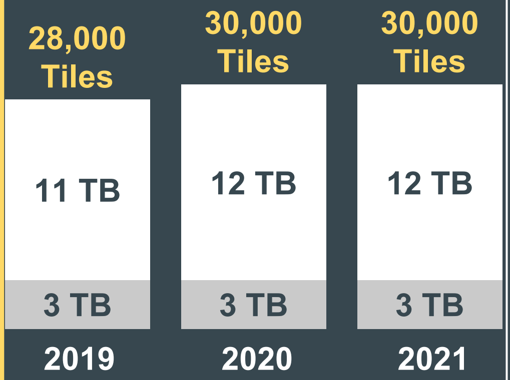
Permits fortnightly monitoring since 2017.

2017-2021 Melt Seasons for both ice sheets:

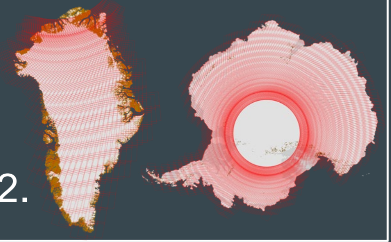
Total data volume = 172 TB

Total tiles = 340,000 Useful Tiles

Landsat-8: Revisit times of ~16 days.



13.



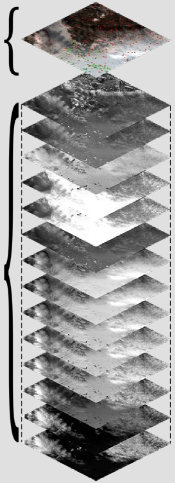
Earlier Landsat since 1972.

Project objective:

Design a workflow which outputs supraglacial hydrology features, from source sensor data, with minimal manual intervention.

Surface-Water and Not-Water points on RGB image.

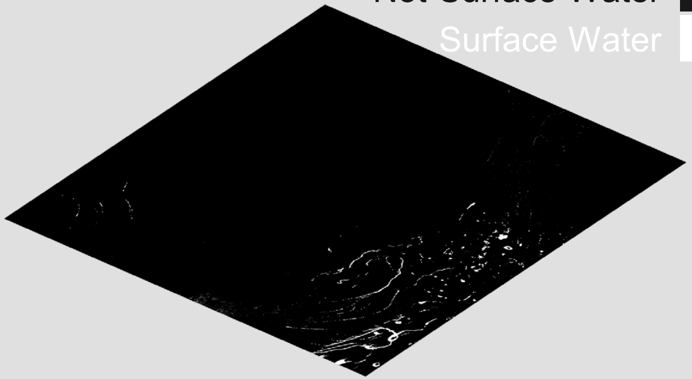
Multi-band tiff: spectral indices + sensor bands.



Apply Algorithm

Not-Surface Water

Surface Water



Binary Surface Water:Not-Surface Water output.

Mapping methods

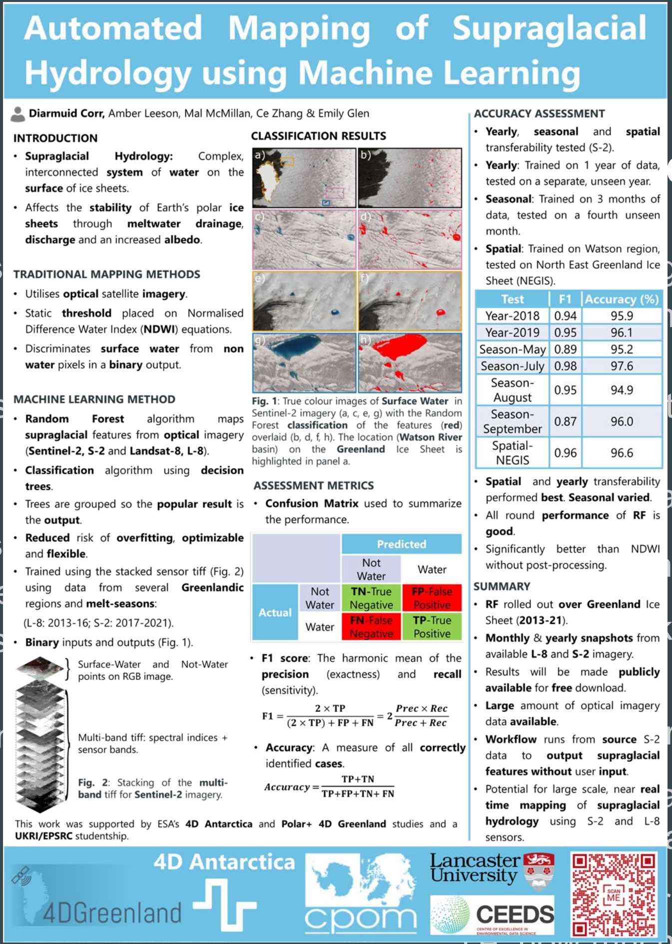
Traditional NDWI

Static thresholds placed on Normalis
Index calculations using optical satell

Results in binary water:not-water clas

Disadvantages:

- Supraglacial hydrology shares
with cloud, shadow, rock & blue
- Results in many false positives
manual post-processing.
- Not feasible for near real-time r



Automated Mapping of Supraglacial Hydrology using Machine Learning

Diarmuid Corr, Amber Leeson, Mal McMillan, Ce Zhang & Emily Glen

INTRODUCTION

- **Supraglacial Hydrology:** Complex, interconnected **system** of **water** on the **surface** of ice sheets.
- Affects the **stability** of Earth's polar **ice sheets** through **meltwater drainage**, **discharge** and an increased **albedo**.

TRADITIONAL MAPPING METHODS

- Utilises **optical** satellite **imagery**.
- Static **threshold** placed on Normalised Difference Water Index (**NDWI**) equations.
- Discriminates **surface water** from **non water** pixels in a **binary** output.

MACHINE LEARNING METHOD

- **Random Forest** algorithm maps **supraglacial** features from **optical** imagery (**Sentinel-2, S-2** and **Landsat-8, L-8**).
- **Classification** algorithm using **decision trees**.
- Trees are grouped so the **popular result** is the **output**.
- **Reduced** risk of **overfitting**, **optimizable** and **flexible**.
- Trained using the stacked sensor tiff (Fig. 2) using data from several **Greenlandic** regions and **melt-seasons**:

(L-8: 2013-16; S-2: 2017-2021).

- **Binary** inputs and outputs (Fig. 1).

Surface-Water and Not-Water points on RGB image.

Multi-band tiff: spectral indices + sensor bands.

Fig. 2: Stacking of the **multi-band** tiff for **Sentinel-2** imagery.

CLASSIFICATION RESULTS

a) b) c) d) e) f) g) h)

Fig. 1: True colour images of **Surface Water** in Sentinel-2 imagery (a, c, e, g) with the Random Forest **classification** of the features (**red**) overlaid (b, d, f, h). The location (**Watson River** basin) on the **Greenland** Ice Sheet is highlighted in panel a.

ASSESSMENT METRICS

- **Confusion Matrix** used to summarize the performance.

		Predicted	
		Not Water	Water
Actual	Not Water	TN-True Negative	FP-False Positive
	Water	FN-False Negative	TP-True Positive

- **F1 score**: The harmonic mean of the **precision** (exactness) and **recall** (sensitivity).

$$F1 = \frac{2 \times TP}{(2 \times TP) + FP + FN} = 2\frac{Prec \times Rec}{Prec + Rec}$$

- **Accuracy**: A measure of all **correctly** identified **cases**.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

ACCURACY ASSESSMENT

- **Yearly**, **seasonal** and **spatial** transferability tested (S-2).
- **Yearly**: Trained on 1 year of data, tested on a separate, unseen year.
- **Seasonal**: Trained on 3 months of data, tested on a fourth unseen month.
- **Spatial**: Trained on Watson region, tested on North East Greenland Ice Sheet (NEGIS).

Test	F1	Accuracy (%)
Year-2018	0.94	95.9
Year-2019	0.95	96.1
Season-May	0.89	95.2
Season-July	0.98	97.6
Season-August	0.95	94.9
Season-September	0.87	96.0
Spatial-NEGIS	0.96	96.6

- **Spatial** and **yearly** transferability performed **best**. **Seasonal varied**.
- All round **performance** of **RF** is **good**.
- Significantly better than NDWI without post-processing.

SUMMARY

- **RF** rolled out **over Greenland** Ice Sheet (**2013-21**).
- **Monthly** & **yearly snapshots** from available **L-8** and **S-2** imagery.
- Results will be made **publicly available** for **free** download.
- **Large** amount of optical imagery **data available**.
- **Workflow** runs from **source** S-2 data to **output supraglacial features without user input**.
- Potential for large scale, near **real time mapping** of **supraglacial hydrology** using S-2 and L-8 sensors.

This work was supported by ESA's **4D Antarctica** and **Polar+ 4D Greenland** studies and a UKRI/EPSRC studentship.

For more info see my poster tomorrow: 17:20-19:00 – Board 70

earning algorithm

gorithm, trained using stacked sensor
n multiple melt-seasons and regions.

thm uses grouped decision trees.

ater:not-water classification.

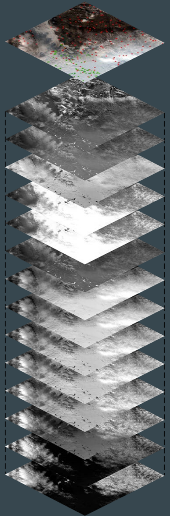
efits:

k of overfitting - separate

rd parameter optimization.

aintains accuracy when a
a is missing.

rmine feature importance -
i.e. how much each band contributes.



Random Forest Classification

The final algorithm was trained on data from multiple regions across 2017-2021 melt-seasons.

Yearly, **seasonal** and **spatial** transferability tested (Sentinel-2).

F1 score: The harmonic mean of the **precision** (exactness) and **recall** (sensitivity).

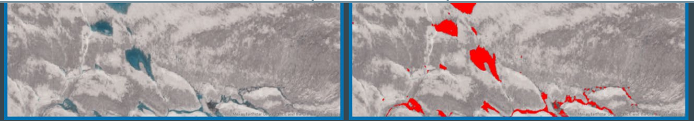
Accuracy: A measure of all **correctly** identified **cases**.

Spatial and **yearly** transferability performed **best**. **Seasonal varied**.

All round **performance** of **RF** is **good**.

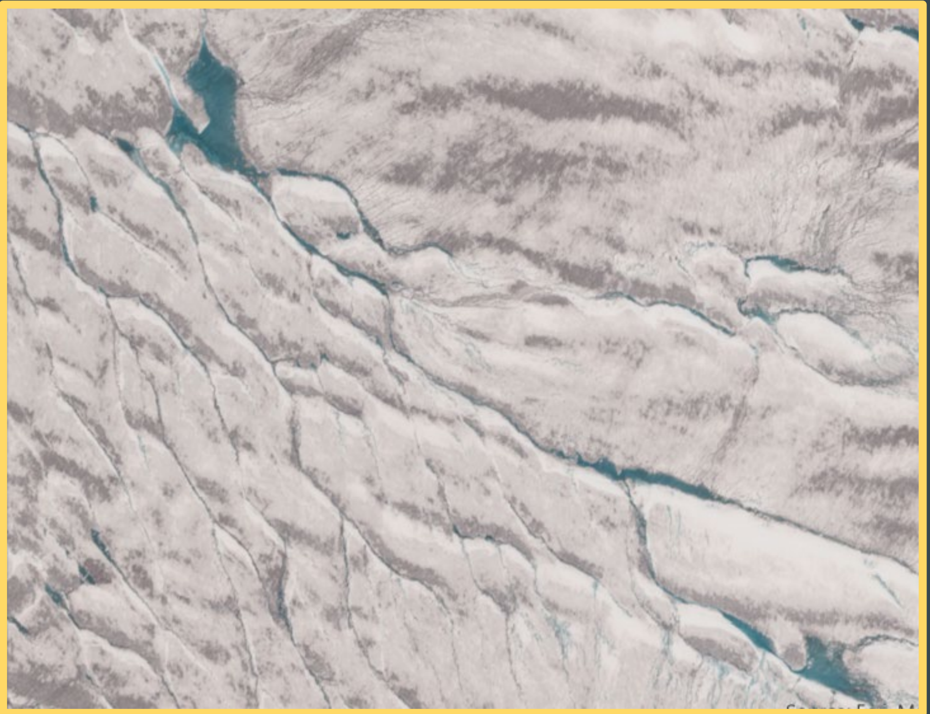
Significantly better than NDWI without post-processing.

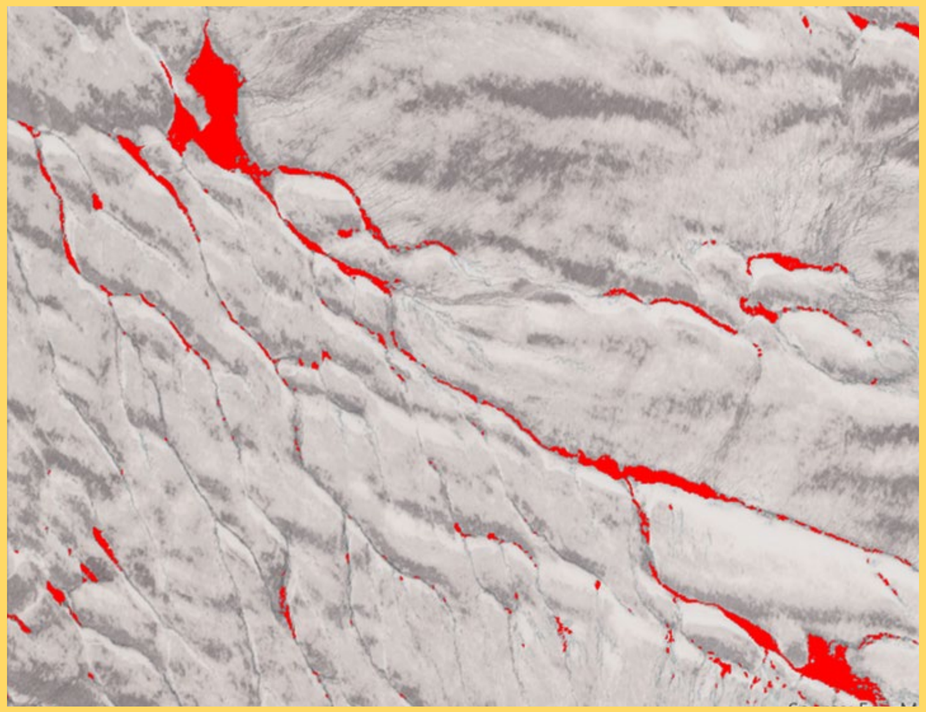
Test	F1	Accuracy (%)
Year-2018	0.94	95.9
Year-2019	0.95	96.1
Season-May	0.89	95.2
Season-July	0.98	97.6
Season-August	0.95	94.9
Season-September	0.87	96.0
Spatial-NEGIS	0.96	96.6



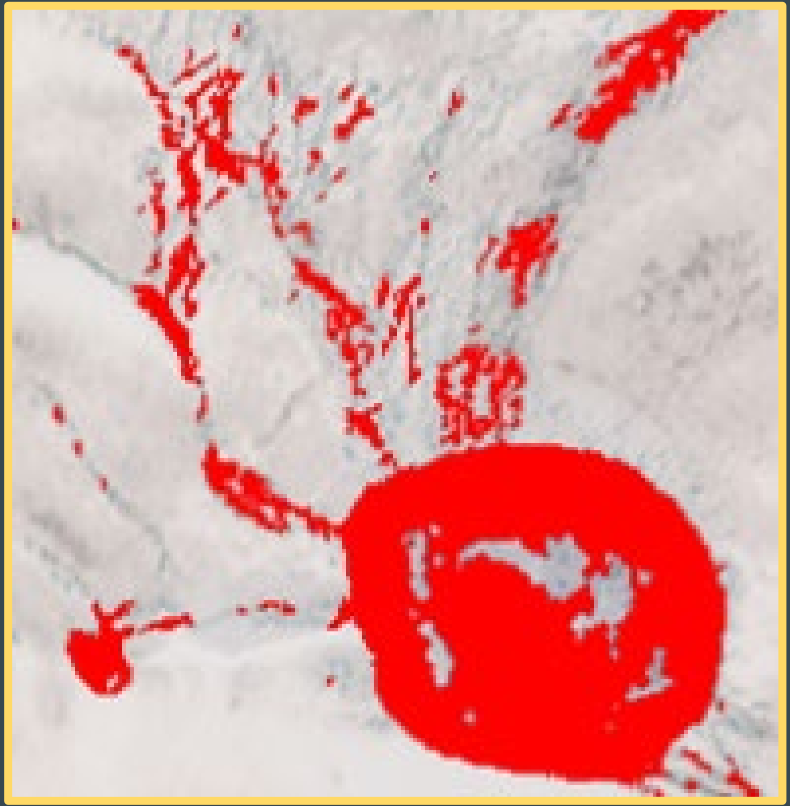
Random Forest Classification - Supraglacial Lake

Random Forest Classification - Supraglacial Channels





Random Forest Classification - Channel-Lake System



Within A Digital Twin

Machine Learning can **automate** historically user-intensive **satellite processing** pipelines.

Important as large **data** quantities and dynamic **features** make manual and traditional **methods ineffective**.

ML has potential for large scale, near **real time mapping** of **supraglacial hydrology**.

Near real time **data** can be **assimilated** into surface hydrological **models**.

Real time data gives a **better understanding** of the hydrological **system**, e.g. used as early warning system for ice shelf collapse in Antarctica.

Results, when used in wider supra and subglacial hydrological models, **may reveal** more about **ice sheet evolution** and the **role** ice sheets play in **ocean systems**.

Within a **Digital Twin**, **models** could **search** for new **patterns** and **features** – e.g. detailed tracking of meltwater dynamics on decadal scales.

Perspectives for Future Work

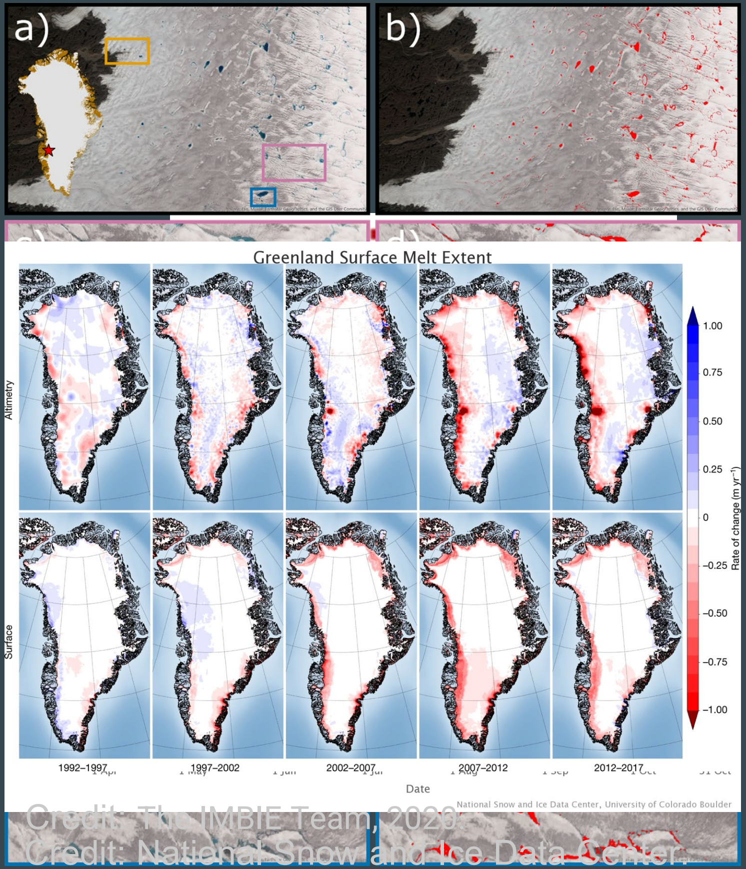
- Link supraglacial lake distribution/depth/volume datasets to process models.
- Extend mapping to older missions (e.g. Landsat-1:7).
- Apply Machine/Deep Learning to map lake depths.
- Incorporate high resolution Digital Elevation Models - Better resolve streams ~10 m.

Takeaways

Supraglacial hydrology is important in a warming world.

ML algorithms can utilise large data quantities to map supraglacial features.

Real-time mapping, within a Digital Twin, can improve understanding of ice sheet evolution.



Lancaster University | CEEDS Centre of Excellence in Environmental Data Science | cpom | 4DGreenland | 4D Antarctica | Digital Twin Antarctica | esa